



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Enabling openness of valuable information resources

Curbing data subtractability and exclusion

Vassilakopoulou, Polyxeni; Skorve, Espen; Aanestad, Margunn

Published in:
Information Systems Journal

DOI (link to publication from Publisher):
[10.1111/isj.12191](https://doi.org/10.1111/isj.12191)

Creative Commons License
CC BY 4.0

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Vassilakopoulou, P., Skorve, E., & Aanestad, M. (2019). Enabling openness of valuable information resources: Curbing data subtractability and exclusion. *Information Systems Journal*, 29(4), 768-786.
<https://doi.org/10.1111/isj.12191>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Enabling openness of valuable information resources: Curbing data subtractability and exclusion

Polyxeni Vassilakopoulou¹  | Espen Skørve² | Margunn Aanestad³

¹University of Agder, Kristiansand, Norway

²Aalborg University, Aalborg, Denmark

³University of Oslo, Oslo, Norway

Correspondence

Vassilakopoulou, Polyxeni, University of Agder, Kristiansand, Norway
Email: polyxenv@uia.no

Funding information

Norges Forskningsråd, Grant/Award Numbers: 210622 and 237766

Abstract

In this paper we investigate how data openness can be made possible in communal settings. We adopt a utility perspective that foregrounds the use value of data, conceptualizing them as “goods.” On the basis of this conceptualization we explore 2 key goods’ attributes: subtractability and exclusion. Our theoretical basis is built upon concepts from the theory of the commons, power theorizing, and notions related to data and information. Empirically, we investigate openness in the genetics domain through a longitudinal study of the evolving communal infrastructure for data related to 2 genes influencing women’s susceptibility to breast and ovarian cancer (*BRCA1* and *BRCA2*). We follow the continuously shifting “topology” of the *BRCA* information infrastructure and trace the multiple repositories that are put in place and the different arrangements for data collection, curation/quality assurance, access, and control that are tried out. In our analysis, we illustrate the actors’ strategies for curbing the subtractability and exclusion attributes of data. We then propose a theoretically informed and empirically grounded framework that can guide understanding and action taking to enable data openness.

KEYWORDS

commons, critical research, open data, power

1 | INTRODUCTION

The advent of genetic testing for clinical purposes and the rapid development of sequencing technologies led to an exponential growth of genetic data. These data are valuable resources for both researchers and clinicians. Researchers

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Information Systems Journal* Published by John Wiley & Sons, Ltd.

draw on them to explore the relationship between genes and health outcomes and clinicians refer to them to support clinical decisions related to diagnostics, treatment planning, and prevention (Battista, Blancquaert, Laberge, Van Schendel, & Leduc, 2012; Bennett, Burke, Burton, & Farndon, 2010; Haga et al., 2012; Skirton, Patch, & Williams, 2005; Snyderman, 2012). Nevertheless, tensions abide in the genetic data governance area and different approaches are being followed for data collection, curation/quality assurance, access and control. There are major initiatives towards data sharing and openness, while there are also significant initiatives that adopt restrictive approaches for pooling together and governing genetic data. The differences among actors' stances towards data openness and enclosure are fueling controversies and contestations.

In this paper, we investigate data openness in the genetics domain through a longitudinal study of the evolving infrastructure for data related to 2 genes that influence women's susceptibility to breast and ovarian cancer (*BRCA1* and *BRCA2*). This is a significant domain as "a large proportion of the work in genetic services is the management of familial breast and ovarian cancer, and this clinical area exemplifies both the opportunities and challenges to increasing access to gene testing" (Slade, Riddell, Turnbull, Hanson, & Rahman, 2015). Our analysis spans the 2 decades elapsed since the identification of the 2 specific genes. We adopt a utility perspective, emphasizing the use value of data and conceptualizing them as "goods." To gain insights, we use the Ostroms' typology of goods (Ostrom & Ostrom, 1977) unpacking the notion of openness.

Our investigative concerns are 3-fold: (1) to bring insights about openness in the specific BRCA data domain, (2) to develop a critique for the prevailing practice of blackboxing openness, and (3) to suggest workable arrangements for facilitating the equitable use of data resources in the field. These investigative concerns position our work within critical information research (Myers & Klein, 2011). Critical information systems research is characterized more by the type of investigative concerns and critical intentions rather than commonality in underlying theories (Brooke, 2002; Ceece-Kecmanovic, Klein, & Brooke, 2008; Stahl, 2008).

Our analysis provides insights about the impediments and opportunities for global data sharing and openness in the much contested BRCA domain. Adopting a utility perspective and conceptualizing data as "goods," we unpack data openness along 2 key goods' attributes: exclusion and subtractability (Ostrom & Ostrom, 1977). Understanding data openness entails examining the processes through which data are generated and used and the processes that make them valuable. We explore these processes in the BRCA data domain by following the evolution of communal BRCA data repositories, and we trace how the subtractability and exclusion attributes of data are continuously negotiated and reshaped. The dynamics of subtractability and exclusion set data apart from biophysical goods where these attributes are exogenous (ie, not shaped within the arena of action). In the contested domain of BRCA data repositories multiple actors leverage different forms of power to pursue or restrict openness by shaping subtractability and exclusion.

The remainder of the paper is structured as follows. First, we present our analytical lens that entails conceptualizing data openness from a utility perspective. Subsequently, we present our research approach including motivation, research context, and methods for data collection and analysis. Then, we present our empirical findings and analysis. Finally, we discuss our findings in light of our overall aims and conclude by presenting implications of our work.

2 | DATA OPENNESS FROM A UTILITY PERSPECTIVE

2.1 | Data as goods

A utility perspective orients attention to the use value of data. This value relates to the information that data can convey to knowledgeable users (Kettinger & Li, 2010). From a utility perspective, data acquire value when used as information resources. The use value of data is contextual and can be identified by tracing the role of data for performing knowledge work tasks and the impact of data on task outputs (Repo, 1986). Hence, this value is dependent upon the arrangements that make possible the realization of their potential (Aaltonen & Tempini, 2014).

Data and information have a recursive relationship: during production, data are created from information (Tuomi, 1999), and during utilization, data convey information to users that have prior knowledge (Kettinger & Li, 2010) and a specific interest (an object) (Zimmerman, 2008). As different users use different knowledge configurations to make sense of data (possibly also having different interests) the same data can result to different information (ie, different object representations) (McKinney & Yoos, 2010).

Three decades ago, Cleveland explored the unique characteristics of information as a resource and noted that it differs from traditional resources in that it is expandable (can increase with use rather than being consumed), compressible, substitutable (in the sense that it can “replace” capital, labor, and physical materials when smartly used), diffusive (it tends to leak), and sharable without being depleted (Cleveland, 1982, 1985). These characteristics of information resources and especially the fact that information is leaky made Cleveland to proclaim that the winds of openness are whistling, signalling the obsolescence of ownership for ideas or facts. Nevertheless, the generic characteristics of information resources are not enough for mandating openness.

Although information is leaky, the rate by which it leaks can be painstakingly slow. In cases where information timeliness and currency is of importance, it is possible to restrict the flow of benefits from information resources by delaying their diffusion. Outdated information resources can bring little or no benefits under time-sensitive conditions. Furthermore, although information resources can be shared without being consumed, in many cases, sharing can diminish the benefits accrued from them. For instance, when information resources are used to reduce capital, labor, or material providing a competitive advantage, there are good reasons for safeguarding them.

To sum-up, approaching data from a utility perspective (conceptualizing them as goods) entails linking data to their use context, thus revealing the diversity of situations that influence their governance. Such a perspective goes beyond the generic characteristics that relate to the nature of information resources (ie, that they are expandable, compressible, substitutable, diffusive, sharable) and brings to the forefront both the characteristics of users (that have specific prior knowledge and interests) and the characteristics of overall arrangements within which information will play a resource role. The specifics of these arrangements are important for exploring the issue of openness.

2.2 | The question of openness: insights from the typology of goods

The question of openness does not arise unless there is the possibility and desirability for non-openness (ie, enclosure). In the late 1970s Ostrom and Ostrom introduced a typology of goods aiming to disambiguate between regimes for openness/enclosure and the nature of goods. For some time, economists had struggled with classifying goods as either private or public. In their typology, the Ostroms clearly identified that there are more than 2 types of goods (Ostrom & Ostrom, 1977). They used 2 attributes from the political economy literature that help identify 4 broad classes (Table 1). The first attribute is about the subtractability of goods' benefits. Subtractability signifies up to what extent the benefits consumed by someone may subtract from the benefits available to others. For instance, in the case of mineral goods like coal and petroleum, the benefits consumed by someone that uses a quantity mined are subtracted from what is available for others. For non-material goods, such as data or information, examples of subtractability include the benefit of scientific publication, where results based on specific data are only novel once. The second attribute relates to how difficult it is to exclude individuals from access to the flow of benefits (Hess & Ostrom, 2003). Mechanisms for exclusion range from regulation of access rights to physical enclosure of the goods or their distribution channels. For instance, exclusion is much easier for arable land than for ocean fisheries.

TABLE 1 Typology of goods based on exclusion potential and subtractability

		Subtractability	
		Low	High
Exclusion	Easy	Club goods	Private goods
	Difficult or infeasible	Public goods	Common-pool resources

Adapted from Ostrom and Ostrom (1977) and Hess and Ostrom (2003).

The critical factor in this approach is to begin with the specifics of the goods involved (Ostrom & Ostrom, 1977). The category of goods for which it is difficult or infeasible to implement exclusion measures for the flow of benefits is characterized as “commons.” The “commons” can be governed as public goods or as common-pool resources. The classification of goods in 1 of the 4 broad classes does not automatically lead to an optimal governance regime but it contours the action arena. For example, common-pool resources can be linked to a variety of regimes: “common-pool resources may be owned by national, regional, or local governments, by communal groups, by private individuals or corporations, or used as open-access resources by whomever can gain access. Each regime has different sets of advantages and disadvantages” (Hess & Ostrom, 2003).

What we found intriguing in the conceptualization proposed by the Ostroms is that the exclusion and subtractability of goods are considered to be “external variables” (ie, exogenous factors not to be shaped within the arena of action). While this might hold for biophysical goods (like ocean fisheries, arable land, forests, and sunshine), it is not always the case with non-material, handmade goods. For this type of goods, the exclusion and subtractability attributes are not necessarily to be taken as given, on the contrary, these characteristics can be fabricated and are technologically contingent. This opens up for an action arena where different actors influence the goods' attributes by exercising different types of power.

2.3 | The role of power within openness-shaping action arenas

Clegg has theorized power within organization fields and proposed 3 circuits through which we can make sense of it: (1) the episodic circuit that is associated to causal power in day to day relations; (2) the dispositional or social integration circuit that is linked to member relations, alliances, and authority legitimization; and (3) the facilitative or system integration circuit that is related to change and the empowerment or disempowerment of agents (Clegg, 1989). What is interesting in this model is that it includes both power exercised episodically (eg, during specific conflict instantiations) and power that shapes the overall action arena (dispositional and facilitative). Furthermore, the model points to the “juncture at which social and system integration meet (...) the stabilization and fixing of rules of meaning and membership, and techniques of production and discipline in an organization field” (Clegg, 1989: 241). Power is not only manifested when conflict is present between the actors that have different stances towards data openness and enclosure. The dispositional (social integration) and facilitative (system integration) circuits are 2 additional types of power exercising that are especially important as actors are actively shaping, rather than simply addressing the exclusion and subtractability properties of data.

3 | RESEARCH MOTIVATION, CONTEXT, AND METHOD

3.1 | Motivation

The impetus for our study comes from our involvement in a research and development project within genetics. This was a collaborative project between the Department of Medical Genetics in a large Scandinavian University Hospital and the University in the same city. The aim of the project was to develop a secure IT platform to facilitate distributed collaboration and access to a high-performance analysis and storage facility. As one of the research activities in that project, we conducted interviews and observations of how molecular biologists and other specialists conduct their work. During our observations we were struck with the role of external databases and tools. We started further investigations into the international “ecology” of data resources via secondary data collection (documents, reports, research papers, etc). Our investigation was focused on BRCA variant assessment datasets, starting with the Breast Information Core, a globally shared database that was established in 1995. We downloaded the content of the database, analyzed the patterns of submissions, and examined the content of the records. This analysis brought to light a surprising finding: although BIC is a highly valuable database for the clinical genetics specialists, the richness of its content decreased

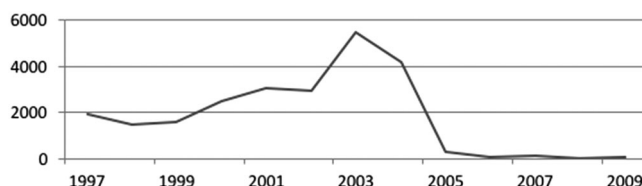


FIGURE 1 Breakdown of Breast Cancer Information Core entries by creation year for the 1997-2009 period (analysis of database instance downloaded on September 17, 2014)

abruptly after 2004 (Figure 1). This was due to the decision of a private company (which has been the most significant contributor to BIC up to that point) to stop sharing data.

This surprising finding triggered our interest to explore issues of openness and enclosure in the domain and led us to investigate the diverse value logics that are present (Vassilakopoulou, Skorve, & Aanestad, 2016a) and to address genetic data governance through a commons perspective (Vassilakopoulou, Skorve, & Aanestad, 2016b). Although our research interest when entering the project was of a technical nature (revolving around software and network solutions that could facilitate practitioners' everyday work), the analysis of the content of the datasets triggered an interest to get better insights on the tensions around data openness within the domain. We realized that to facilitate the unobstructed and equitable use of valuable data resources in the domain, it was not sufficient to propose technical solutions (for user-friendly interface design and for efficient linkages among heterogeneous data containers). It was also important to take a critical perspective and aim for insights to address longstanding patterns of exclusion.

Pursuing this research interest, we studied the evolution of BIC starting from its inception and following its trajectory looking at key decision points and conflicts over the years. Starting from BIC we expanded our coverage to other related initiatives and collated data to produce an account of the information landscape evolution with the aim of examining how openness is achieved or obstructed. On the basis of a critical approach, we understand the domain of data sharing as a contested domain and data as valuable goods around which contests may arise.

3.2 | Research context: BRCA testing

In 1990, the geneticist Mary-Claire King and her team discovered the linkage between a gene on chromosome 17 and breast cancer predisposition (Hall et al., 1990). The discovery of this linkage triggered an intense race among research teams around the world to sequence the gene and to develop relevant genetic testing procedures that led to the identification of *BRCA1* in 1994 and of *BRCA2* in 1995. Having a pathogenic *BRCA1* or *BRCA2* variant can increase a woman's risk of developing breast cancer to between 50% and 80% (the general population risk is 12%) and the risk of developing ovarian cancer to between 24% and 40% (the general population risk is 1%-2%) (Petrucelli, Daly, & Feldman, 2013). This means that a BRCA test can be used not only for diagnostic purposes but also as a pre-symptomatic test to check for disease predisposition. The identification of the BRCA genes made genetic testing possible not only for diseases that are inherited from generation to generation (like thalassemia or cystic fibrosis) but also for diseases where genes are one of many causative factors (eg, both inheritance and environment play a role).

Although a predisposition does not mean that the disease will certainly appear (significant numbers of women that carry pathogenic BRCA mutations have reached their 90s without developing breast or ovarian cancer), a positive test can trigger medical interventions that can be as radical as mastectomy or ovary removal. Genetic testing in the 2 BRCA genes is aimed towards identifying the presence or absence of cancer predisposing gene variants (mutations). Testing entails mapping the gene sequence for a specific individual and comparing it with what is most commonly encountered in the general population. Differentiations (variants) from the common sequence are assessed by experts and classified as follows: variants that indicate pathogenicity, variants that do not indicate pathogenicity, or variants of uncertain/unclassified clinical significance. The assessment of variants' pathogenicity is based on experts'

interpretations. The quality of these interpretations is highly dependent on the evaluation of prior evidence, and for this reason, data repositories play a critical role (Aronson & Rehm, 2015).

The process of BRCA genetic testing includes sampling (eg, taking a blood sample), sample processing with the use of a specialized machine (sequencer), variant identification using informatics tools, and variant assessment. Each part of the process is significant. Nevertheless, performing the first stages of the process that lead to variant identification is relatively straightforward and largely automated. The major challenges remain for variant assessment (Quintáns, Ordóñez-Ugalde, Cacheiro, Carracedo, & Sobrido, 2014). *BRCA1* and *BRCA2* are genes that produce tumour suppressor proteins. These proteins help repair damaged DNA. When either of these genes is altered in way that its protein product is not made or does not function correctly, DNA damage may not be repaired properly. As a result, cells are more likely to develop additional genetic alterations that can lead to cancer (National Cancer Institute at the National Institute of Health—USA, 2014). Since not all variants have the same impact on gene function, the identification of a variant (ie, a differentiation from what is most commonly encountered in the general population) is not enough to signify cancer predisposition. The impact of the variant has to be assessed before reaching a conclusion.

To perform the assessment, experts search for prior variant classifications in datasets containing anonymized results from past BRCA genetic tests. Search can be performed in local data repositories (containing results from past tests performed by the same laboratory) or in shared ones (where many laboratory test results deposit test results) or in published scientific papers and reports. Additionally to looking up prior classifications, scientists can perform different types of analysis for the variant identified. Such analysis may entail specialized predictive models and software packages. It may also entail experimentation and measurements performed on organic material (as opposed to computer simulations), in a controlled environment to define protein functionality. Furthermore, family studies (performing tests to relatives) may also be initiated. The availability of good quality datasets on past assessments can significantly facilitate the assessment of variants minimizing the need for further analysis and reducing the duplication of efforts within laboratories around the world. In this context, the role of databases is critical. The laboratories that have access to reliable data on past assessments can expedite their work, while laboratories that have no such access (and the population that gets tested in these laboratories) are disadvantaged. In that sense, the issue of openness and enclosure in the domain is of significant societal relevance.

3.3 | Method for data collection and analysis

In the previous sections we presented the research context and the dynamics of our research process. We provided an account of our research motives explaining how they were guided by the unfolding pattern of findings (McGrath, 2005). Our concern for investigating openness in the BRCA domain oriented our attention to the interests that are driving enclosure and the arrangements in place that dictate the inclusion or exclusion from the flow of data benefits. For our investigation we followed a process approach paying attention on how and why things emerge, develop, grow, or terminate over time (Langley, Smallman, Tsoukas, & Van de Ven, 2013). Our methodological approach has similarities to genealogical approaches adopted by Foucault; historical investigations into the events that have led to the current status. Foucault eschews grand narratives and emphasizes the importance of the local relations and actions (Klecut, 2004), and this paradigm can guide research that aspires to focus on the negotiated nature of established arrangements. To build our empirical basis, we aimed to (re)construct the BRCA data repositories' trajectory that spans more than 2 decades by collecting documentary evidence, performing semistructured interviews and inspecting digital artefacts (actual BRCA repositories). This effort yielded a significant volume of data (Table 2).

We analyzed the data collected by focusing on the evolution of events related to BRCA data governance. To gain insights, we used the Ostroms' typology of goods (Ostrom & Ostrom, 1977) for unpacking the notion of openness in the domain. Initially, the collected data were analyzed by reconstructing a chronology of events. Then, we analyzed our empirical material by putting together episodes that relate to subtractability and exclusion. As we proceeded with the analysis, we identified gaps in our empirical material and we extended the data collection. We also engaged in

TABLE 2 Data sources

Source	Description	Topics Covered
Interviews	12 semistructured interviews with scientists engaged in medical genetics (biologists, bioinformaticians, medical doctors).	Repositories used and work practices, data exchanges with other scientists insights on ongoing developments in the field related to data sharing. These interviews helped us to orient within the field and to make sense of data utility.
Data repositories inspection	Inspected the BIC, ClinVar and BRCA exchange repositories	For BIC: analyzed the patterns of submissions and examined the content of the records. For ClinVar: retrieved BRCA variant data and examined the information consolidation mechanisms used. For BRCA Exchange: retrieved BRCA variant data and explored the different views provided (including a curated list of BRCA variants, interpreted by expert consensus).
Document analysis	Journal papers, media publications, scientific guidelines, nomenclature documents, legal proceedings, project reports, presentations prepared for various audiences	Reviewed 330 documents to identify key events that mark the evolution of the overall information infrastructure and to gather information on public debates for controversial issues.

discussions with domain experts, exposing our findings. In several instances, domain experts would direct us to relevant additional sources or point out relevant episodes. Our analysis brings insights about the exclusion and subtractability attributes of gene-specific datasets of genetic variant assessments. For such valuable sociotechnical goods, the attributes that shape openness are contingent to the overall arrangements they relate. These arrangements are in flux as technology is being leveraged simultaneously by several actors to enable favourable conditions for openness and by several other actors to create favourable conditions for enclosure.

4 | A BRIEF GENEALOGY OF COMMUNAL REPOSITORIES FOR BRCA VARIANTS

In the sections that follow we present a brief genealogy of communal repositories for data on BRCA variants. Our narrative includes a part dedicated to open communal repositories and a part on non-open hybrid communal repositories with restrictions. The narrative is complemented by a table that provides an overview of both types of communal BRCA data repositories (Table 3). In the figure that follows (Figure 2) we provide a timeline of key milestones for the different initiatives described. The milestones are colour-coded; white boxes illustrate initiatives towards open data sharing while grey boxes illustrate initiatives that adopt restrictive approaches.

4.1 | The advent of open communal repositories

The need for putting in place communal variant repositories was identified early after the BRCA discoveries. In 1995, 10 scientists from universities, hospitals, a research institute, and a private company in different European countries and the United States created an open Web-accessible repository for BRCA data named Breast Cancer Information Core (BIC) (Friend et al., 1995). BIC's aim is explained in a 1996 paper: "One of the serious impediments to achieving clinical benefits from the isolation of the BRCA1 gene is finding and assessing the significance of mutations in this new cancer susceptibility gene. This will be greatly facilitated by coordinated detection and interpretation of mutations and

TABLE 3 Communal BRCA data repositories as of April 2017

Repository Name	When Started	Deposit Policy	Access Policy	BRCA Variants	Description
BIC	1995	Open discretionary curated	Open after registration	4 000	A Web-accessible repository that facilitates the accumulation and sharing of BRCA variant assessments. Data are accessible by all. New data included after being reviewed by volunteering experts.
HGMD	1996	Curated pulled from literature	Updated data by subscription. 3-year old data open after registration.	3 900	A repository for gene variants and their interpretation accessible via subscription. A free version is also provided but contains data appearing with a time delay. The variants listed in are manually curated from the scientific literature.
LOVD shared	2005	Open discretionary curated (some "pull" mechanisms in place)	Open	3 900	A free, open software for the creation of locus-specific databases developed by the Leiden University. Database administrators who have set up their own installation can activate the "share option." Publicly available Web interface.
SCRIP	2012	Open discretionary	Open (via ClinVar)	2 000	A mechanism for collecting BRCA variant assessments via doctors that receive test results from Myriad in the United States.
ClinVar	2012	Open discretionary	Open	12 400	An aggregator of variant interpretations (in an open, Web-accessible repository). The repository covers all genes (not only BRCA variants). Data deposited by multiple contributors—not curated.
BRCA Exchange	2015	Pulled from multiple sources, curated	Open	16 900	A Web portal to share expert-reviewed BRCA variant classifications. A new pull model is implemented for the proactive identification and aggregation of data.
BRCA Mutation Database	2015	Curated pulled from literature	Open	2 600	An open repository realized after a University of Utah initiative. Entries are created "in-house" through extensive literature curation and the application of computational methods.
BRCA Share	2015 (based on UMD established in 2000)	Participants only (full sharing required), curated	Participants only	6 200	A public-private data sharing initiative. Participation is possible for all laboratory test results on a fee basis (free for research) as long as they share all their data.

the dissemination of this information to as many qualified investigators as possible. To this end, the BIC has created and maintains a central repository for information regarding mutations and polymorphism" (Couch & Weber, 1996).

BIC has accumulated up to today more than 30 000 anonymized entries corresponding to approximately 4000 distinct *BRCA1* and *BRCA2* variants. Data are deposited by individual investigators, research, hospital-based, and commercial laboratory test results, and are published after being examined by BIC's steering committee members. Registration is open to all, and access to registered users is unrestricted. Hundreds of submitters around the world support

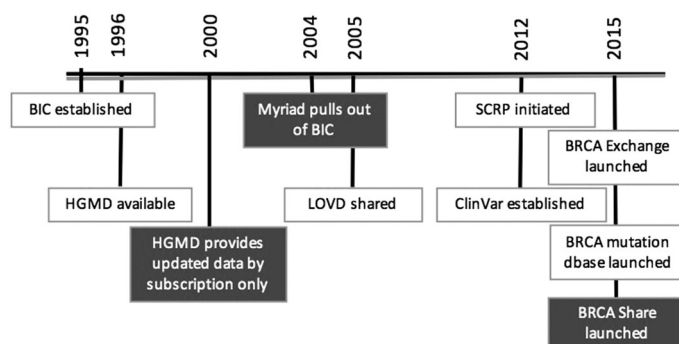


FIGURE 2 Timeline of communal BRCA repositories (white boxes illustrate initiatives for open data sharing, and grey boxes illustrate initiatives for restrictive approaches)

BIC, making it one of the most rich and comprehensive information resources available. Nevertheless, when taking into account the growth in BRCA testing, it becomes evident that the repository receives only a small portion of the globally generated information. For instance, it has been reported that more than 100 000 BRCA tests are performed annually in the United States alone (Armstrong et al., 2015).

Myriad is the most publicized non-contributor to BIC. This is the world's largest molecular diagnostic clinical laboratory (Myriad Genetics, 2014), which dominates the BRCA genetic testing market in the United States. Myriad was founded in 1991 as a spin-off from the University of Utah's Centre for Genetic Epidemiology to support the geneticist Mark Skolnick and his team in the sequencing of *BRCA1*, and in the subsequent development of a test. Skolnick's team won the *BRCA1* race against many other research groups and filed patent applications on *BRCA1*, and soon after on *BRCA2*. Myriad's dominant position in genetic testing can be attributed to the 9 *BRCA1* and *BRCA2* related patents, which the company was granted with between 1997 and 2000. Myriad was the most significant contributor to BIC until 2004 when a decision to stop contributing was taken. The company explained this decision by claiming that BIC lacks operational and clinical standards and includes data of questionable quality as a high proportion of variants is classified as of "unknown clinical importance" (Angrist & Cook-Deegan, 2014; Tucker, 2014). In 2013, the US Supreme Court invalidated Myriad's patents but the company retained a competitive advantage by having accumulated a unique dataset in a proprietary database (Baldwin & Cook-Deegan, 2013). The database is being used internally for looking-up variants traced in the past and has also been leveraged for the development of a scoring algorithm for variant assessment (Pruss et al., 2014).

As a reaction to Myriad's decision to discontinue BIC contributions, 2 geneticists initiated a project named "Sharing Clinical Reports Project" (SCRCP) reaching out for physicians that receive Myriad's test reports (Nguyen & Terry, 2013). SCRCP was initiated in 2012 (Kolata, 2013) and received since then more than 5000 submissions related to more than 2000 unique BRCA variants (Sharing Clinical Reports Project, 2014). Soon after the initiation of SCRCP a complementary project entitled "Free the Data" was launched. This project addresses patients providing them a platform for sharing their own BRCA reports and giving them the option to contribute additional health-related information (Free the Data, 2014).

In 2012, the US National Center for Biotechnology Information released a new open repository (named ClinVar) to host anonymized data on variant interpretations (Landrum et al., 2013) and invited scientific groups that hold data from clinical testing, research, and literature to contribute their information. ClinVar is repository that covers all human gene variations and is not dedicated only to the BRCA genes. A number of hospitals and laboratory test results that were not submitting actively to BIC started submitting data to ClinVar. Additionally, both BIC and SCRCP are registered ClinVar submitters and have contributed interpretations for all their variants. Submitted interpretations for a particular variant are presented through comprehensive views that associate multiple submissions. Conflict resolutions can be provided from expert panels or professional societies (Landrum et al., 2013). Unlike BIC where the content of submissions is reviewed by members of its steering committee and each variant interpretation is periodically

updated to reflect the latest findings, in ClinVar the original content of the submissions is not curated or modified. As of April 2017, ClinVar hosted information on 12 400 *BRCA1* and *BRCA2* variants (among them 842 with conflicting interpretations). One of the positive aspects of ClinVar, according to the medical director of the molecular genetics and genomics department of a large laboratory, is that it alerts laboratory test results when the classification for a variant differs between them: "We can get on the phone and talk with the other laboratory. We may simply have different internal evidences that we are using" (Ray, 2015). By sharing information laboratory test results can advance their understanding on *BRCA* variants and improve their interpretations.

Both BIC and ClinVar rely on discretionary submissions of multiple laboratories that "push" information in the shared repository. Following a different, "pull" approach, in June 2015 a new Web portal named BRCA Exchange was launched under the auspices of the Global Alliance for Genomics and Health. The Global Alliance is an impressive network of major health care providers, research funders, research institutions, disease advocacy organizations, and life science and information technology companies. The network's aim is to create a common framework of harmonized approaches to enable the responsible, voluntary, and secure sharing of genomic and clinical data (Global Alliance for Genomics and Health, 2015). BRCA Exchange aims to facilitate open sharing of expert-reviewed *BRCA* variant classifications. A number of working groups with members located around the world are working within this initiative to identify sources of classified variants to be aggregated and also, unclassified variants with evidence, which can then be interpreted, classified, and aggregated. Essentially, BRCA Exchange does not rely to the depositors' initiative for data collection but rather, proactively exploits existing data sources through standardized application programming interfaces (APIs) and procedural conventions. ClinVar is used as a major source of information. The classified variants are displayed at the BRCA Exchange portal for "use by clinicians, patients, and labs to enable accurate understanding of any individual *BRCA1/2* variant, for superior clinical care and decision-making" (BRCA Exchange, 2015). As of April 2017, BRCA Exchange included more than 16 900 unique *BRCA* variants.

Following another approach for pulling together existing information and making it easily available to all, the University of Utah's Department of Pathology together with its nonprofit ARUP Laboratories launched in May 2015 the open *BRCA* Mutation Database (Ray, 2015) in collaboration with the University's Huntsman Cancer Institute and with the WHO International Agency for Research on Cancer. This new open communal repository for information on *BRCA1* and *BRCA2* variants provides 2 types of datasets: (1) classified variants found in the literature and family studies (containing information on more than 2600 variants) and (2) in silico risk predictions for all possible single nucleotide substitutions in *BRCA1* and *BRCA2* to support variant interpretation. Practically, the *BRCA* Mutation Database is not populated by information deposited by third parties or pulled from pre-existing databases but the entries are created "in-house" through extensive literature curation and the application of computational methods.

4.2 | The advent of hybrid communal repositories with restrictions

During the 1990s a team within the French Institute of Health and Medical Research (INSERM) initiated the development of a generic software called UMD (Universal Mutation Database) for the creation of locus-specific databases. Locus-specific databases are collections of variants limited to a specific gene (eg, the *BRCA* genes), as opposed to more general databases. The first version of UMD was released in 2000, and since then it has been continuously upgraded and improved (Bérout et al., 2005; Bérout, Collod-Bérout, Boileau, Soussi, & Junien, 2000; Caputo et al., 2012). UMD was used to develop *BRCA1* and *BRCA2* databases that include curated, compiled information from 16 laboratories located all over France belonging to the GGC consortium (Groupe Génétique et Cancer). These databases have been endorsed by the French National Cancer Institute (INCa) and were designed to collect all *BRCA* variants detected in France.

In April 2015, INSERM in collaboration with the commercial laboratory Quest Diagnostics announced the launch of BRCA Share, which is a novel *BRCA* repository that builds upon the existing *BRCA* data in the UMD *BRCA* databases (the repository contains data on more than 6200 *BRCA* variants). BRCA Share is a public-private initiative, with

Quest licensing the data and forming sublicense agreements with participants. To participate in BRCA Share, laboratories have to commit on sharing past, present, and future data. Commercial laboratory test results need to pay a subscription fee according to their size, while research entities get access at no cost. This financial arrangement allows BRCA Share to invest on data curation arrangements to attend to data quality and to conduct functional studies on the effects of mutations, without depending on research or public funding. Curation refers to the quality assurance process of data and often includes manual evaluations by multiple specialists. This is a crucial issue that requires extensive resources, especially when the depositors and their local procedures prior to submission are not well known. Essentially, BRCA Share's alternative model resembles a "club arrangement" and was described by a Nature editorial as a "walled garden" (Nature Editorial, 2015).

A similar logic to the one introduced by UMD has been followed for the Leiden Open Variation Database (LOVD), which is a dedicated software for the creation of locus-specific databases developed by the Leiden University Medical Center in the Netherlands. LOVD is a free, open-source tool that was launched in 2005 and is actively being improved, currently having new releases every month (Fokkema et al., 2011; Fokkema, den Dunnen, & Taschner, 2005). Database administrators who have set up their own installation can activate the "share option" for anonymized data. The Web interface is publicly available and can be freely searched, but other activities, including sequence variant submission, require prior registration. The LOVD shared installation (which is accessible by anybody) includes more than 3900 unique public BRCA variants.

Back in 1996, a repository named Human Gene Mutation Database (HGMD) was made publicly available (Cooper & Krawczak, 1996; Krawczak et al., 2000). The repository was maintained at the Institute of Medical Genetics in Cardiff in the United Kingdom and contained data on gene mutations underlying human inherited disease. HGMD was built upon a database that 2 geneticists developed in the early 1990s to collate data for the study of mutation mechanisms in human genes (Krawczak et al., 2000; Krawczak & Cooper, 1995). In 2000, HGMD entered a licensing agreement with the company Celera Genomics and agreed to provide Celera exclusive access to new information added for a 1-year period (Stenson et al., 2003). The updated HGMD content was thereafter available on a paid subscription basis, while a free version was also provided via the Cardiff website (containing data appearing with a time delay). This agreement lasted until 2005. Since 2005, HGMD has been working in partnership with another company (BIOBASE GmbH) using a similar business model (Stenson et al., 2012). The publicly available version of HGMD is still maintained and is accessible free of charge but is out-of-date by 3 years. Since 2006, access to the publicly available version is provided after registration and is only granted to users from academic/nonprofit institutions (Stenson et al., 2009). The variants listed in HGMD are sourced and manually curated from the scientific literature. Relevant literature is identified via a combination of manual journal screening and automated procedures covering more than 1950 journals (Stenson et al., 2014). HGMD does not include only data related to *BRCA1* and *BRCA2* but its BRCA related content is rich: the free version of HGMD contains more than 3000 BRCA variants (the subscription version contains about 30% more).

4.3 | Current state: a fragmented landscape

The initiatives presented in the previous sections are summarized in Table 3 revealing the fragmentation and heterogeneity of the evolving BRCA information infrastructure.

The multiplicity of repositories does not only result to duplication of efforts and increased difficulty in data retrieval. More significantly, as different repositories follow different approaches for data collection and curation they end up conveying inconsistent information. In a comparative study (initiated by Myriad), the data content across 5 different BRCA data repositories (BIC, ClinVar, HGMD Pro, LOVD, and UMD) were analyzed (Vail et al., 2015). The researchers investigated how many of a set of 1327 variants from Myriad's database were included in the communal repositories (only 124 were found in all 5) and compared the classifications for internal and cross-repository consistency. The finding of this investigation was that different sources often provide different assessments for the same variant.

5 | ANALYSIS

5.1 | Data flows within the domain

The continuous efforts of multiple distributed actors make possible the flow of BRCA data globally. Patient samples are processed in genetic laboratories where the variants identified are assessed and filed in local repositories. The assessment work is supported by data on past variant assessments retrieved from communal repositories. Some of the newly generated data are shared (submitted for publication to journals or deposited for inclusion in repositories). Data curators and repository managers are facilitating and controlling data gathering (overview in Figure 3).

The sharing of variant assessments is pivotal for distributing the labor required for characterizing variants and avoiding the duplication of efforts, for identifying and resolving differences in interpretation, and for providing a catalogue of variation that could be used for research. However, the accumulated data in the communal repositories are only a small part of the data generated in the various laboratories. Large volumes of data are stored locally and remain inaccessible as significant effort is required for putting data in a shareable format and linking them or uploading them to communal repositories. In some cases, laboratories may also have an incentive for keeping their data for proprietary use, especially in the cases when they have to operate within a market environment. Even data that become part of the public domain (eg, through scientific publications) require significant effort and costs for becoming easily retrievable. Furthermore, some of the repositories that pool together this type of data are only available on a subscription basis. Although several scientists have been vocal about the need to ensure sharing in the domain (Cook-Deegan, Conley, Evans, & Vorhaus, 2013; Field et al., 2009; Matloff, Barnett, & Nussbaum, 2014; Nature Editorial, 2015; Tucker, 2014) and large-scale investments supported by governments and international organizations are made, universal data sharing remains a vision. This situation is linked to the exclusion and subtractability attributes of BRCA datasets that do not foster a universal approach to openness. In the following sections we explore the shaping of those 2 key attributes.

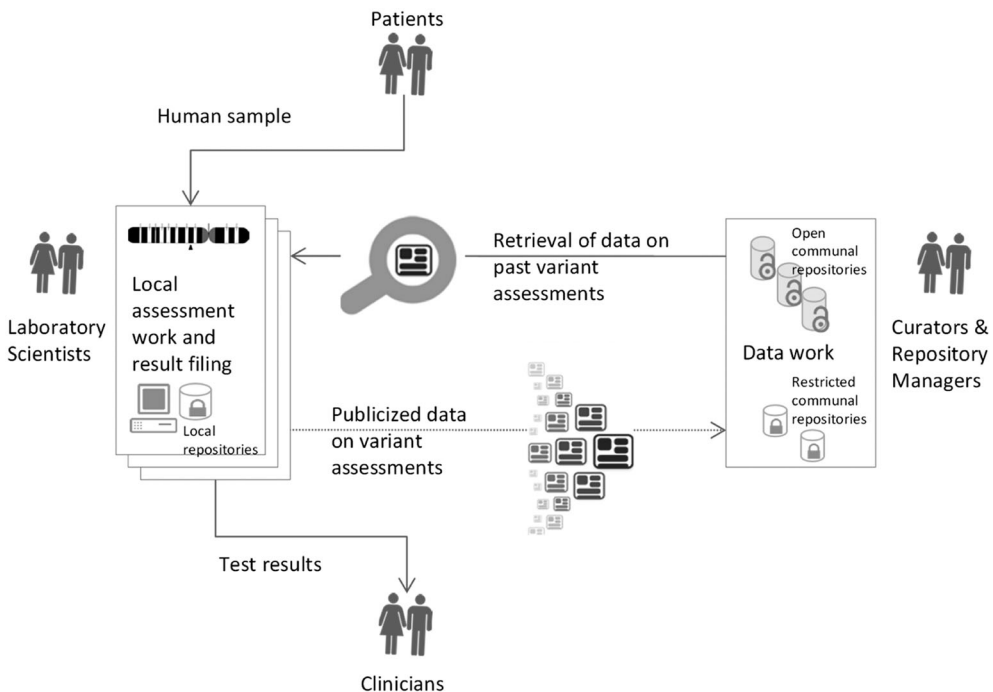


FIGURE 3 Data flows within the BRCA information infrastructure

5.2 | Subtractability

In the case of private laboratories that compete within market conditions data sharing can be undesirable. Actually, the benefits that other laboratories might appropriate by reusing the data shared can subtract to an extent from the advantage of the laboratory that completed the assessment in the first place. Through data sharing, laboratories that have already incurred assessment costs enable competitors to benefit without exerting the same efforts, losing some competitive advantage when the same variant is encountered. Actually, 2 of the unique characteristics of information as a resource, substitutability and expandability (Cleveland, 1982, 1985), are creating the conditions for high subtractability. Information on past variant assessments can substitute some of the analysis labor required creating a competitive advantage for those that possess such information. Moreover, information on variants is expandable; it can be further manipulated with advanced models to produce new insights that commercial laboratories may exploit as trade secrets.

To counteract the tendency of laboratories to keep valuable information for their own use, practice guidelines mandate public disclosure when “a variant of uncertain clinical significance becomes clearly pathogenic, or a variant is not pathogenic anymore” (Wallis et al., 2013). Furthermore, proponents of openness are negotiating with the College of American Pathologists (a leading accreditation organization for diagnostic laboratories) to make data sharing a condition of membership, while some of the insurance companies consider making it a requirement for reimbursement (Krol, 2014).

5.3 | Exclusion

As BRCA data are generated in multiple laboratories around the world, they are first registered in multiple dispersed sites. The variant assessments are communicated to health care providers and patients; therefore, practically it is not possible to prevent individuals from disseminating them as the SCRP initiative proves. Furthermore, exclusion is considered incongruent with the scientific tradition of medical genetics, where data sharing is valued not only for the potentiality of advancing scientific knowledge but also for the preservation of information and for safeguarding against misconduct (van Schaik, Kovalevskaya, Protopapas, Wahid, & Nielsen, 2014). Still, exclusion phenomena prevail.

Information as a resource is diffusive and sharable (Cleveland, 1982, 1985). Technology offers the possibility for information to be easily digitized and shared, and it can start flowing towards public databases but this kind of flow can be painstakingly slow. Standardization and networking have been used by different actors in the BRCA field for making sharing a seamless activity and accelerating data flows. For instance, in the case of laboratories that use internally the same software (eg, LOVD or UMD) there is the possibility of real-time data sharing. Pull mechanisms to proactively amass data have also been created (as in the case of BRCA Exchange) to ensure more timely and comprehensive data sharing. Furthermore, specialized tools to facilitate data evaluation during curation activities are being used to significantly reduce the effort required for ensuring data quality. Overall, connectivity and communality (Fulk, Flanagan, Kalman, Monge, & Ryan, 1996) are being used by several key actors to counteract exclusion tendencies in the domain. Connectivity is pursued via technical solutions and standards, while communality is pursued via curators, trusted third parties, or specialized committees.

6 | DISCUSSION

In our analysis we explored how data openness can be unpacked by conceptualizing data as valuable goods and by leveraging the typology of goods proposed by Ostrom and Ostrom (1977). This typology is built around 2 key attributes: subtractability and exclusion. While contributions on data openness abound in the IS literature, including studies on open data in health care and public sector (eg, Alanazi & Chatfield, 2012; Kuk & Davies, 2011; Laine, Lee, & Nieminen, 2015), it is striking that the meaning of data openness is black-boxed. We suggest that unpacking openness

is important for contextualizing and operationalizing the concept, and the typology offered by Ostrom and Ostrom is one possible approach towards achieving this. As the analysis of the evolution of data repositories in the BRCA domain shows, subtractability and exclusion are being shaped by actors that participate in the field with different positions and different interests. The genetics field is densely populated by public and private actors of varying sizes, different roles (eg, laboratories performing tests, information service providers, clinical units, research units), and different business models. The exclusion and subtractability of data goods are not exogenous factors (as for biophysical goods) but rather are being fabricated by the different actors within this arena of action. Subtractability and exclusion are shaped by all these participants (that can be data depositors, data users, data curators) and also by actors that finance health care (that set their own rules having an interest in improving efficiency), health policy makers (that take measures aiming to the sustainability and the continuous improvement of health systems), and citizens (that influence the political arena). Ensuring access to data is important for pursuing human well-being, for complying with norms of scientific evidence, and for advancing medicine, but is not yet built into the system (Cook-Deegan & McGuire, 2017). This has been emphasized by the National Research Council of the US National Academies that identified the need to build "information commons" based on data sharing to advance precision medicine and improve health (National Research Council, 2011).

Many groups have been working during the past 2 decades to address BRCA data sharing issues, pooling data on BRCA genetic variants sourced from laboratories, research enterprises, clinicians, and published research through diverse approaches and data governance schemes (Bérout et al., 2000; Cooper & Krawczak, 1996; Fokkema et al., 2005; Friend et al., 1995; Landrum et al., 2013; Lawler et al., 2015). Despite the intense activity in the domain, the various efforts have not converged into a dominant governance model and the vision of ensuring access to data remains largely unfulfilled. Our analysis shows that developing workable arrangements for openness requires finding ways to ensure that the benefits consumed by data users will not subtract from the benefits of actors that possessed the data in the first place. Furthermore, it requires finding ways to prevent exclusion from access to the flow of benefits. This points to the importance of power exercising through the dispositional and facilitative circuits. The dispositional (social integration) circuit provides the conditions for actors to exercise power. The importance of this dispositional power for pursuing openness or enclosure (depending on actors' interests) is manifested in efforts to promote data sharing in policy documents, and guidelines by openness proponents. At the same time, those that oppose openness are developing arguments based on the (lack of) data quality in open repositories (Vail et al., 2015). While dispositional power is concerned with the capacities to configure legitimization conditions, facilitative power can be both disciplinary and generative as it relates to the creation of outcomes and change. It is defined by the techniques of production and discipline within the domain. It relates to systems of rewards and punishment and technologies, job design, and networks. Facilitative power is manifested in the BRCA case as new technologies are leveraged within targeted initiatives that aim to disrupt the status in the domain (eg, SCRP and BRCA Exchange). Clegg uses the term "techniques of production" for the technological means of controlling the physical and social environment in organizations and refers to these techniques as an example of facilitative power that comes into play in the system integration circuit (Clegg, 1989). The role of information and communication technologies for facilitating openness in the domain is key. New governance arrangements will require thinking of actors' relations by taking into account also technology design (Van den Hoven, 2007). The impressive growth of the BRCA Exchange dataset is related to the decision to leverage technologies for automatically pulling data from multiple pre-existing repositories. Furthermore, the connectivity afforded by Web technologies is leveraged for enrolling multiple qualified scientists from all over the world in massive curation and quality assurance activities. The BRCA Exchange initiative targeted simultaneously data sourcing, quality assurance, and access.

Power and divergent interests are ever present in domains as complex as the genetics one. The traditional Weberian definition of power ("the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance, regardless of the basis on which this probability rests" (Weber, 1947: 152)) is often interpreted as precluding manifestations of power unless there is some sort of resistance (or conflict). Nevertheless, power exercised through the dispositional and systemic circuits is key for establishing workable arrangements for

TABLE 4 Possible interventions to influence exclusion and subtractability

		Possible Interventions to Shift Towards Openness	
		Dispositional Power	Facilitative Power
Influence on data (goods) attributes	Exclusion	For instance, building a case for sharing in policy documents, building relations and alliances.	For instance, through technologies that facilitate data pulls and distributed curation.
	Subtractability	For instance, working with social norms for contributions within collectives, building a case against free-riding.	For instance, through technologies to trace and attribute contributions, mechanisms for credit to parties involved, rules for test provider selection.

openness. In Table 4, we present possible ways to shift the domain towards openness by influencing the subtractability and exclusion attributes of BRCA data. These ways entail leveraging dispositional and facilitative power and are built upon the findings from the case analyzed. In a complex global domain as the one studied, there will always be multiple actors working under different jurisdictions, with different interests and orientations. Hence, although it is difficult for openness to be mandated, the conditions for openness can be cultivated in the domain by those interested in advancing medical knowledge and human well-being. Unpacking openness and making sense of how different forms of power exercising can shape the action arena is a good start for such interventions.

Interventions like the ones presented in Table 4 can support the creation of a more favourable to openness environment. This would not preclude individual actors from pursuing strategies of enclosure. As Cook-Deegan and McGuire have noted in a recent paper, the viability of the openness model depends in part on how good public databases become; if communal data sources accessible to all “catch up” to proprietary ones, strategies based on enclosure will lose much of their marginal value (Cook-Deegan & McGuire, 2017). In this context, we especially want to draw attention to the role of information systems as means for exercising facilitative power to shape the subtractability and exclusion attributes in the BRCA data domain. There are already significant successes with technological arrangements that facilitate data flows (addressing the exclusion characteristic of BRCA data). A next step could be to capitalize on technological possibilities for enhancing non-subtractability. This can be aimed, for instance, through technologies that trace and attribute contributions, and mechanisms for credit to parties involved. Technology can also be used to strengthen strategies of enclosure either within private or club arrangements. Information systems research can contribute in the shaping of the domain, revealing and explaining both the role of information and communication technologies for producing socially desirable consequences and the risks of building technology-enabled enclosure arrangements.

7 | CONCLUSION

Sharing valuable data at a global scale depends on the sustained work of many participants. In this paper we examined the issue of openness by analyzing a specific case of genetic data (anonymized datasets on BRCA variant assessments). Our analysis can provide a better understanding of the impediments to global sharing and the opportunities for moving towards more open arrangements. Understanding data openness entails examining the processes through which data are generated and used and the processes that make them valuable. We explored these processes in the BRCA data domain by following the evolution of communal repositories showing how the subtractability and exclusion attributes of BRCA data are not exogenously defined but are continuously negotiated and reshaped. Following our investigative concerns we bring insights about openness in the specific domain and we contribute to unpacking openness through conceptualizing BRCA data as goods that can be analyzed in terms of subtractability and exclusion using the typology proposed by the Ostroms (Ostrom & Ostrom, 1977). Furthermore, we suggest ways to put in place workable arrangements for promoting data openness in the field by pointing to the role of non-causal power. Having an action orientation is an implication of taking a critical stance: “the responsibility of a researcher in a social situation

does not end with the development of sound explanations and understandings of it" (Ngwenyama & Lee, 1997: 151). Specifically, we identify how dispositional and facilitative power (Clegg, 1989) can be leveraged to shift the domain towards openness by curbing the subtractability and exclusion attributes of data. Hence, we contribute a theoretically informed and empirically grounded framework that can guide both the assessment of the data sharing situation in the field and action taking. Although the framework is illustrated by interventions (Table 4) that are specific to the BRCA data domain and the multiple systems and practices within its information infrastructure (Figure 3), it can be re-contextualized for analyzing different data domains and identifying interventions.

Taking a utility perspective on data (conceptualizing them as goods) and thinking in terms of dispositional and facilitative power can guide understanding and action taking. Within the current dynamic and multi-actor data landscapes, it is possible to take openness-oriented initiatives that are different from mandating openness (through causal power). Dispositional and facilitative power can be exercised to curb exclusion and subtractability by targeted interventions in data sourcing (mobilizing and facilitating depositing), data quality assurance (stimulating and supporting curation activities), data access (promoting and easing search and retrieval), and data crediting-rewarding (fostering and enabling tracing, attribution, and rewarding of contributions). Such initiatives can be part of concerted action for data openness in domains where the societal significance of openness is justified. Relevant domains include scientific research (OECD, 2004) and public government (Dawes, Vidasova, & Parkhimovich, 2016). Information systems research offers a unique knowledge base that can advance open data across domains (Link et al., 2017). Openness can be a specific praxis-oriented theme for critical information systems research aiming to more equitable data arrangements (Richardson & Robinson, 2007).

ACKNOWLEDGEMENTS

This research was supported by the Research Council of Norway (Norges Forskningsråd) projects nr. 210622 and 237766.

ORCID

Polyxeni Vassilakopoulou  <http://orcid.org/0000-0002-5947-4070>

REFERENCES

- Aaltonen, A., & Tempini, N. (2014). Everything counts in large amounts: A critical realist case study on data-based production. *Journal of Information Technology*, 29(1), 97–110.
- Alanazi, J. and Chatfield, A. (2012), *Sharing government-owned data with the public: A cross-country analysis of open data practice in the Middle East*, AMCIS Proceedings.
- Angrist, M., & Cook-Deegan, R. (2014). Distributing the future: The weak justifications for keeping human genomic databases secret and the challenges and opportunities in reverse engineering them. *Applied & Translational Genomics*, 3(4), 124–127.
- Armstrong, J., Toscano, M., Kotchko, N., Friedman, S., Schwartz, M., Virgo, K., ... Bauer, J. (2015). Utilization and outcomes of BRCA genetic testing and counseling in a national commercially insured population: The ABOUT study. *JAMA Oncology*, 1–10.
- Aronson, S., & Rehm, H. (2015). Building the foundation for genomics in precision medicine. *Nature*, 526(7573), 336–342.
- Baldwin, A. L., & Cook-Deegan, R. (2013). Constructing narratives of heroism and villainy: Case study of Myriad's BRACAnalysis® compared to Genentech's Herceptin®. *Genome Medicine*, 5(1), 8.
- Battista, R. N., Blancquaert, I., Laberge, A.-M., Van Schendel, N., & Leduc, N. (2012). Genetics in health care: An overview of current and emerging models. *Public Health Genomics*, 15(1), 34–45.
- Bennett, C., Burke, S., Burton, H., & Farndon, P. (2010). A toolkit for incorporating genetics into mainstream medical services: Learning from service development pilots in England. *BMC Health Services Research*, 10(1), 125.
- Bérout, C., Collod-Bérout, G., Boileau, C., Soussi, T., & Junien, C. (2000). UMD (Universal Mutation Database): A generic software to build and analyze locus-specific databases. *Human Mutation*, 15(1), 86–94.
- Bérout, C., Hamroun, D., Collod-Bérout, G., Boileau, C., Soussi, T., & Claustres, M. (2005). UMD (Universal Mutation Database): 2005 update. *Human Mutation*, 26(3), 184–191.

- BRCA Exchange (2015), History of the BRCA Exchange. <<http://brcaexchange.org/about/history%3e>, accessed 15 October 2015.
- Brooke, C. (2002). What does it mean to be 'critical' in IS research? *Journal of Information Technology*, 17(2), 49–57.
- Caputo, S., Benboudjema, L., Sinilnikova, O., Rouleau, E., Bérout, C., Lidereau, R. and the French Brca GGC Consortium (2012), Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases, *Nucleic Acids Research*, 40 (Database issue), D992-D1002.
- Cecez-Kecmanovic, D., Klein, H. K., & Brooke, C. (2008). Exploring the critical agenda in information systems research. *Information Systems Journal*, 18(2), 123–135.
- Clegg, S. R. (1989). *Frameworks of power*. London: SAGE Publications.
- Cleveland, H. (1982). Information as a resource. *Futurist*, 16(6), 34–39.
- Cleveland, H. (1985). The twilight of hierarchy: Speculations on the global information society. *Public Administration Review*, 45(1), 185–195.
- Cook-Deegan, R., Conley, J., Evans, J., & Vorhaus, D. (2013). The next controversy in genetic testing: Clinical data as trade secrets? *European Journal of Human Genetics*, 21(6), 585–588.
- Cook-Deegan, R., & McGuire, A. L. (2017). Moving beyond Bermuda: Sharing data to build a medical information commons. *Genome Research*, 27(6), 897–890. <https://doi.org/10.1101/gr.%20216911.116>
- Cooper, D., & Krawczak, M. (1996). Human Gene Mutation Database. *Human Genetics*, 98(5), 629–629.
- Couch, F., & Weber, B. L. (1996). Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. *Human Mutation*, 8(1), 8–18.
- Dawes, S., Vidasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15–27.
- Field, D., Sansone, S., Collis, A., Booth, T., Dukes, P., Gregurick, S., Kennedy, K., Kolar, P., Kolker, E., and Maxon, M. (2009), *Omics data sharing, science* (New York, NY), 326 (5950), 234.
- Fokkema, I., den Dunnen, J., & Taschner, P. (2005). LOVD: Easy creation of a locus-specific sequence variation database using an "LSDb-in-a-box" approach. *Human Mutation*, 26(2), 63–68.
- Fokkema, I., Taschner, P., Schaafsma, G., Celli, J., Laros, J., & den Dunnen, J. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, 32(5), 557–563.
- Free the Data (2014), Genetic information is more valuable when shared. <www.free-the-data.org> accessed 15 September 2014.
- Friend, S., Borresen, A. L., Brody, L., Casey, G., Devilee, P., Gayther, S., ... Wiseman, R. (1995). Breast cancer information on the web. *Nature Genetics*, 11(3), 238–239.
- Fulk, J., Flanagan, A., Kalman, M., Monge, P., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory*, 6(1), 60–87.
- Global Alliance for Genomics and Health (2015), About the global alliance. <<https://genomicsandhealth.org/about-the-global-alliance/frequently-asked-questions%3e>, accessed 26 October 2015.
- Haga, S. B., O'Daniel, J. M., Tindall, G. M., Mills, R., Lipkus, I. M., & Agans, R. (2012). Survey of genetic counselors and clinical geneticists' use and attitudes toward pharmacogenetic testing. *Clinical Genetics*, 82(2), 115–120.
- Hall, J., Lee, M., Newman, B., Morrow, J., Anderson, L., Huey, B., & King, M. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 250(4988), 1684–1689.
- Hess, C., & Ostrom, E. (2003). Ideas, artifacts, and facilities: Information as a common-pool resource. *Law and Contemporary Problems*, 66(1/2), 111–145.
- Kettinger, W., & Li, Y. (2010). The infological equation extended: Towards conceptual clarity in the relationship between data, information and knowledge. *European Journal of Information Systems*, 19(4), 409–421.
- Klecun, E. (2004). Conducting critical research in information systems: Can actor-network theory help? *Information systems research: Springer*, 259–274.
- Kolata, G. (2013), DNA project aims to make public a company's data on cancer genes, The New York Times, (April 12 2013). <http%3A%2F%2Fwww.nytimes.com%2F2013%2F04%2F13%2Fhealth%2Fdna-project-aims-to-make-companys-data-public.html%3Fpagewanted%3Dall%26amp%3B_r%3D0%253e%0A, accessed 15 September 2014.
- Krawczak, M., Ball, E., Fenton, I., Stenson, P., Abeyasinghe, S., Thomas, N., & Cooper, D. (2000). Human Gene Mutation Database—A biomedical information and research resource. *Human Mutation*, 15(1), 45–51.
- Krawczak, M., & Cooper, D. (1995). Core database. *Nature*, 374(6521), 402–402.

- Krol, A. (2014). As genetics moves to the clinic, pathogenic variants still subject to doubt and debate, BIO IT World. <<http://www.bio-itworld.com/2014/4/17/genetics-moves-clinic-pathogenic-variants-still-subject-doubt-debate.html%3e>, accessed December 31, 2015.
- Kuk, G. and Davies, T. (2011). The roles of agency and artifacts in assembling open data complementarities, ICIS 2011 Proceedings.
- Laine, S., Lee, C., and Nieminen, M. (2015). *Transparent data supply for open information production processes*, Proceedings of the European Conference on Information Systems.
- Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2013). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*. gkt1113
- Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, 56(1), 1–13.
- Lawler, M., Siu, L., Rehm, H., Chanock, S., Alterovitz, G., Burn, J., ... North, K. (2015). All the world's a stage: Facilitating discovery science and improved cancer care through the Global Alliance for Genomics and Health. *Cancer Discovery*, 5(11), 1133–1136.
- Link, G., Lombard, K., Conboy, K., Feldman, M., Feller, J., George, J., ... Willis, M. (2017). Contemporary issues of open data in information systems research: Considerations and recommendations. *Communications of the Association for Information Systems*, 41(1), 587–610.
- Matloff, E., Barnett, R., & Nussbaum, R. (2014). Choosing a BRCA genetic testing laboratory: A Patient-centric and ethical call to action for clinicians and payers evidence-based oncology. 20 (7 (May 2014), SP229–SP230.
- McGrath, K. (2005). Doing critical research in information systems: A case of theory and practice not informing each other. *Information Systems Journal*, 15(2), 85–101.
- McKinney, E., & Yoos, C. (2010). Information about information: A taxonomy of views. *MIS Quarterly*, 34(2), 329–344.
- Myers, M. D., & Klein, H. K. (2011). A set of principles for conducting critical research in information systems. *MIS Quarterly*, 35(1), 17–36.
- Myriad Genetics (2014) Myriad genetics corporate presentation, <<https://myriad-web.s3.amazonaws.com/myriad.com/pdf/Myriad-Corporate-Presentation.pdf>>
- National Cancer Institute at the National Institute of Health—USA Titl', (updated 22 Januaray 2014) <<http://www.cancer.gov/cancertopics/factsheet/Risk/BRCA%3e>, accessed 07 August 2014.
- National Research Council (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. National Academies Press.
- Nature Editorial (2015). Thank you for sharing. *Nature*, 520(7549), 585.
- Nguyen, S., & Terry, S. F. (2013). Free the data: The end of genetic data as trade secrets. *Genetic Testing and Molecular Biomarkers*, 17(8), 579–580.
- Ngwenyama, O., & Lee, A. (1997). Communication richness in electronic mail: Critical social theory and the contextuality of meaning. *MIS Quarterly*, 21(2), 145–167.
- OECD (2004). Science, technology, and innovation for the 21st century, meeting of the OECD committee for science and technological policy at ministerial level, Paris, 29 to 30 January 2004. <<http://www.oecd.org/science/sci-tech/sciencetechnologyandinnovationforthe21stcenturymeetingoftheoecdcommitteeforscientificandtechnologicpolicyatministeriallevel29-30january2004-finalcommunique.htm%3e>, accessed 20 November 2017.
- Ostrom, V., & Ostrom, E. (1977). Public goods and public choices. In E. S. Savas (Ed.), *Alternatives for delivering public services: Toward improved performance*.
- Petrucelli, N., Daly, M. B., & Feldman, G. L. (2013). BRCA1 and BRCA2 hereditary breast and ovarian cancer. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, H. C. Mefford, et al. (Eds.), *GeneReviews*. Seattle: University of Washington.
- Pruss, D., Morris, B., Hughes, E., Egginton, J. M., Esterling, L., Robinson, B. S., ... Gutin, A. (2014). Development and validation of a new algorithm for the reclassification of genetic variants identified in the BRCA1 and BRCA2 genes. *Breast Cancer Research and Treatment*, 147(1), 119–132.
- Quintáns, B., Ordóñez-Ugalde, A., Cacheiro, P., Carracedo, A., & Sobrido, M. J. (2014). Medical genomics: The intricate path from genetic variant identification to clinical interpretation. *Applied & Translational Genomics*, 3(3), 60–67.
- Ray, T. (2015). ARUP takes on BRCA variant classification by Launching Database, in silico VUS resource, GenomeWeb, (May 13, 2015). <<https://www.genomeweb.com/business-news/arup-takes-brca-variant-classification-launching-database-silico-vus-resource%3e>, accessed 28 December 2015.

- Repo, A. (1986). The dual approach to the value of information: An appraisal of use and exchange values. *Information processing & management*, 22(5), 373–383.
- Richardson, H., & Robinson, B. (2007). The mysterious case of the missing paradigm: A review of critical information systems research 1991–2001. *Information Systems Journal*, 17(3), 251–270.
- Sharing Clinical Reports Project (2014), Sharing clinical reports project. <<http://www.iccg.org/about-the-iccg/collaborations/sharing-clinical-reports-project/%3e>, accessed 15 September 2014.
- Skirton, H., Patch, C., & Williams, J. (2005). *Applied genetics in healthcare: A handbook for specialist practitioners*. Garland Science.
- Slade, I., Riddell, D., Turnbull, C., Hanson, H., & Rahman, N. (2015). Development of cancer genetic services in the UK: A national consultation. *Genome Medicine*, 7(1), 18.
- Snyderman, R. (2012). Personalized health care: From theory to practice. *Biotechnology Journal*, 7(8), 973–979.
- Stahl, B. C. (2008). The ethical nature of critical research in information systems. *Information Systems Journal*, 18(2), 137–163.
- Stenson, P., Ball, E., Mort, M., Phillips, A., Shaw, K., & Cooper, D. (2012). *The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution*. Current Protocols in Bioinformatics: John Wiley & Sons, Inc.
- Stenson, P., Ball, E., Mort, M., Phillips, A., Shiel, J., Thomas, N., ... Cooper, D. (2003). Human Gene Mutation Database (HGMD®): 2003 update. *Human Mutation*, 21(6), 577–581.
- Stenson, P., Mort, M., Ball, E., Howells, K., Phillips, A., Thomas, N., & Cooper, D. (2009). The Human Gene Mutation Database: 2008 update. *Genome Medicine*, 1(1), 1–6.
- Stenson, P., Mort, M., Ball, E., Shaw, K., Phillips, A., & Cooper, D. (2014). The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics*, 133(1), 1–9.
- Tucker, K.I. (2014), 'Genetics lab refuses to share data that could save Lives', Jewish daily forward, 12 August 2014.
- Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems*, 16(3), 103–117.
- Vail, P., Morris, B., van Kan, A., Burdett, B., Moyes, K., Theisen, A., ... Egginton, J. (2015). Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases. *Journal of community genetics*, (epub ahead of print, 1–9.
- Van den Hoven, J. (2007). ICT and value sensitive design. In P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa, & V. Laurent (Eds.), *The Information Society: Innovations, Legitimacy, Ethics and Democracy* (Vol. 233) IFIP International Federation for Information Processing (pp. 67–72). Boston: Springer.
- van Schaik, T. A., Kovalevskaya, N. V., Protopapas, E., Wahid, H., & Nielsen, F. G. G. (2014). The need to redefine genomic data sharing: A focus on data accessibility. *Applied & Translational Genomics*, 3(4), 100–104.
- Vassilakopoulou, P., Skorve, E., & Aanstad, M. (2016a). Premises for clinical genetics data governance: Grappling with diverse value logics. *The Ethics of Biomedical Big Data: Springer*, 239–256.
- Vassilakopoulou, P., Skorve, E., & Aanstad, M. (2016b). A commons perspective on genetic data governance: The case of BRCA data, European conference on information systems (ECIS). *Research Paper*, 136.
- Wallis, Y., Payne, S., McNulty, C., Bodman, D., Sistermans, E., Robertson, K., Moore, D., Abbs, S., Deans, Z., and Devereau, A. (2013), Practice guidelines for the evaluation of pathogenicity and the reporting of sequence variants in clinical molecular genetics, association for clinical genetic science of the British Society for Genetic Medicine. <http://www.acgs.uk.com/media/774853/evaluation_and_reporting_of_sequence_variants_bpgs_june_2013_-_finalpdf.pdf>, accessed 8 November 2014.
- Weber, M. (1947). *The theory of social and economic organization*. New York: Oxford University Press.
- Zimmerman, A. S. (2008). New knowledge from old data the role of standards in the sharing and reuse of ecological data, science. *Technology & Human Values*, 33(5), 631–652.

How to cite this article: Vassilakopoulou P, Skorve E, Aanstad M. Enabling openness of valuable information resources: Curbing data subtractability and exclusion. *Info Systems J.* 2019;29:768–786. <https://doi.org/10.1111/isj.12191>